

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/646,470A  
Source: 1Fw/6  
Date Processed by STIC: 7/17/06

# ***ENTERED***



IFW16

## RAW SEQUENCE LISTING

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:06

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw

3 <110> APPLICANT: TAKEDA SAN DIEGO, INC.  
 5 <120> TITLE OF INVENTION: CRYSTALLIZATION OF CATHEPSIN S  
 7 <130> FILE REFERENCE: CATS-5002-C1  
 9 <140> CURRENT APPLICATION NUMBER: 10/646,470A  
 10 <141> CURRENT FILING DATE: 2003-08-22  
 12 <150> PRIOR APPLICATION NUMBER: US 60/405,423  
 13 <151> PRIOR FILING DATE: 2002-08-23  
 15 <160> NUMBER OF SEQ ID NOS: 4  
 17 <170> SOFTWARE: PatentIn version 3.3  
 19 <210> SEQ ID NO: 1  
 20 <211> LENGTH: 331  
 21 <212> TYPE: PRT  
 22 <213> ORGANISM: Homo sapiens  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: MISC\_FEATURE  
 27 <222> LOCATION: (1)..(331)  
 28 <223> OTHER INFORMATION: Amino acid sequence for full-length human wild type Cathepsin

S

30 <300> PUBLICATION INFORMATION:  
 31 <308> DATABASE ACCESSION NO: AF230097  
 32 <309> DATABASE ENTRY DATE: 2002-04-08  
 33 <313> RELEVANT RESIDUES: (1)..(331)  
 35 <400> SEQUENCE: 1  
 37 Met Lys Arg Leu Val Cys Val Leu Leu Val Cys Ser Ser Ala Val Ala  
 38 1 5 10 15  
 41 Gln Leu His Lys Asp Pro Thr Leu Asp His His Trp His Leu Trp Lys  
 42 20 25 30  
 45 Lys Thr Tyr Gly Lys Gln Tyr Lys Glu Lys Asn Glu Glu Ala Val Arg  
 46 35 40 45  
 49 Arg Leu Ile Trp Glu Lys Asn Leu Lys Phe Val Met Leu His Asn Leu  
 50 50 55 60  
 53 Glu His Ser Met Gly Met His Ser Tyr Asp Leu Gly Met Asn His Leu  
 54 65 70 75 80  
 57 Gly Asp Met Thr Ser Glu Glu Val Met Ser Leu Met Ser Ser Leu Arg  
 58 85 90 95  
 61 Val Pro Ser Gln Trp Gln Arg Asn Ile Thr Tyr Lys Ser Asn Pro Asn  
 62 100 105 110  
 65 Arg Ile Leu Pro Asp Ser Val Asp Trp Arg Glu Lys Gly Cys Val Thr  
 66 115 120 125  
 69 Glu Val Lys Tyr Gln Gly Ser Cys Gly Ala Cys Trp Ala Phe Ser Ala  
 70 130 135 140  
 73 Val Gly Ala Leu Glu Ala Gln Leu Lys Leu Lys Thr Gly Lys Leu Val  
 74 145 150 155 160  
 77 Ser Leu Ser Ala Gln Asn Leu Val Asp Cys Ser Thr Glu Lys Tyr Gly

## RAW SEQUENCE LISTING

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:06

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw

```

78          165          170          175
81 Asn Lys Gly Cys Asn Gly Gly Phe Met Thr Thr Ala Phe Gln Tyr Ile
82          180          185          190
85 Ile Asp Asn Lys Gly Ile Asp Ser Asp Ala Ser Tyr Pro Tyr Lys Ala
86          195          200          205
89 Met Asp Leu Lys Cys Gln Tyr Asp Ser Lys Tyr Arg Ala Ala Thr Cys
90          210          215          220
93 Ser Lys Tyr Thr Glu Leu Pro Tyr Gly Arg Glu Asp Val Leu Lys Glu
94 225          230          235          240
97 Ala Val Ala Asn Lys Gly Pro Val Ser Val Gly Val Asp Ala Arg His
98          245          250          255
101 Pro Ser Phe Phe Leu Tyr Arg Ser Gly Val Tyr Tyr Glu Pro Ser Cys
102          260          265          270
105 Thr Gln Asn Val Asn His Gly Val Leu Val Val Gly Tyr Gly Asp Leu
106          275          280          285
109 Asn Gly Lys Glu Tyr Trp Leu Val Lys Asn Ser Trp Gly His Asn Phe
110          290          295          300
113 Gly Glu Glu Gly Tyr Ile Arg Met Ala Arg Asn Lys Gly Asn His Cys
114 305          310          315          320
117 Gly Ile Ala Ser Phe Pro Ser Tyr Pro Glu Ile
118          325          330

```

121 &lt;210&gt; SEQ ID NO: 2

122 &lt;211&gt; LENGTH: 996

123 &lt;212&gt; TYPE: DNA

124 &lt;213&gt; ORGANISM: Homo sapiens

127 &lt;220&gt; FEATURE:

128 &lt;221&gt; NAME/KEY: misc\_feature

129 &lt;222&gt; LOCATION: (1)..(996)

130 &lt;223&gt; OTHER INFORMATION: Human cDNA sequence for Cathepsin S

132 &lt;400&gt; SEQUENCE: 2

```

133 atgaaacggc tggtttgtgt gctcttggtg tgctcctctg cagtggcaca gttgcataaa      60
135 gatacctacc tggatcacca ctggcatctc tggaagaaaa cctatggcaa acaatacaag      120
137 gaaaagaatg aagaagcagt acgacgtctc atctgggaaa agaatactaaa gtttgtgatg      180
139 cttcacaacc tggagcattc aatgggaatg cactcatatg atctgggcat gaaccacctg      240
141 ggagacatga ccagtgaaga agtgatgtct ttgatgagtt ccctgagagt tcccagccag      300
143 tggcagagaa atatcacata taagtcaaac cctaatacga tattgcctga ttctgtggac      360
145 tggagagaga aaggggtgtg tactgaagtg aaatatcaag gttcttgtgg tgcttgctgg      420
147 gcttttcagt ctgtgggggc cctggaagca cagctgaagc tgaaaacagg aaagctgggtg      480
149 tctctcagtg cccagaacct ggtggattgc tcaactgaaa aatatggaaa caaaggctgc      540
151 aatggtggct tcatgacaac ggctttccag tacatcattg ataacaaggg catcgactca      600
153 gacgcttcct atccctacaa agccatggat ctgaaatgtc aatatgactc aaaatatcgt      660
155 gctgccacat gttcaaagta cactgaactt ctttatggca gagaagatgt cctgaaagaa      720
157 gctgtggcca ataaaggccc agtgtctgtt ggtgtagatg cgcgtcatcc ttctttcttc      780
159 ctctacagaa gtggtgtcta ctatgaacca tctgtactc agaatgtgaa tcatgggtgta      840
161 cttgtggttg gctatggtga tcttaatggg aaagaatact ggcttgtgaa aaacagctgg      900
163 ggccacaact ttggtgaaga aggatatatt cggatggcaa gaaataaagg aaatcattgt      960
165 gggattgcta gctttccctc ttaccagaaa atctag

```

168 &lt;210&gt; SEQ ID NO: 3

169 &lt;211&gt; LENGTH: 225

## RAW SEQUENCE LISTING

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:06

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw

170 <212> TYPE: PRT  
 171 <213> ORGANISM: Artificial  
 173 <220> FEATURE:  
 174 <223> OTHER INFORMATION: Amino acid sequence for residues 114-331 of Cathepsin S with

a

175 C-terminal Glycine-6x-histidine tag  
 178 <220> FEATURE:  
 179 <221> NAME/KEY: MISC\_FEATURE  
 180 <222> LOCATION: (1)..(218)  
 181 <223> OTHER INFORMATION: Amino acid sequence for residues 114-331 of Cathepsin S  
 183 <220> FEATURE:  
 184 <221> NAME/KEY: MISC\_FEATURE  
 185 <222> LOCATION: (219)..(225)  
 186 <223> OTHER INFORMATION: Amino acid sequence for C-terminal Glycine-6x-histidine tag  
 188 <400> SEQUENCE: 3  
 190 Ile Leu Pro Asp Ser Val Asp Trp Arg Glu Lys Gly Cys Val Thr Glu  
 191 1 5 10 15  
 194 Val Lys Tyr Gln Gly Ser Cys Gly Ala Cys Trp Ala Phe Ser Ala Val  
 195 20 25 30  
 198 Gly Ala Leu Glu Ala Gln Leu Lys Thr Gly Lys Leu Val Ser  
 199 35 40 45  
 202 Leu Ser Ala Gln Asn Leu Val Asp Cys Ser Thr Glu Lys Tyr Gly Asn  
 203 50 55 60  
 206 Lys Gly Cys Asn Gly Gly Phe Met Thr Thr Ala Phe Gln Tyr Ile Ile  
 207 65 70 75 80  
 210 Asp Asn Lys Gly Ile Asp Ser Asp Ala Ser Tyr Pro Tyr Lys Ala Met  
 211 85 90 95  
 214 Asp Gln Lys Cys Gln Tyr Asp Ser Lys Tyr Arg Ala Ala Thr Cys Ser  
 215 100 105 110  
 218 Lys Tyr Thr Glu Leu Pro Tyr Gly Arg Glu Asp Val Leu Lys Glu Ala  
 219 115 120 125  
 222 Val Ala Asn Lys Gly Pro Val Ser Val Gly Val Asp Ala Arg His Pro  
 223 130 135 140  
 226 Ser Phe Phe Leu Tyr Arg Ser Gly Val Tyr Tyr Glu Pro Ser Cys Thr  
 227 145 150 155 160  
 230 Gln Asn Val Asn His Gly Val Leu Val Val Gly Tyr Gly Asp Leu Asn  
 231 165 170 175  
 234 Gly Lys Glu Tyr Trp Leu Val Lys Asn Ser Trp Gly His Asn Phe Gly  
 235 180 185 190  
 238 Glu Glu Gly Tyr Ile Arg Met Ala Arg Asn Lys Gly Asn His Cys Gly  
 239 195 200 205  
 242 Ile Ala Ser Phe Pro Ser Tyr Pro Glu Ile Gly His His His His His  
 243 210 215 220  
 246 His  
 247 225  
 250 <210> SEQ ID NO: 4  
 251 <211> LENGTH: 340  
 252 <212> TYPE: PRT  
 253 <213> ORGANISM: Artificial  
 255 <220> FEATURE:

## RAW SEQUENCE LISTING

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:06

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw

256 <223> OTHER INFORMATION: Amino acid sequence for Cathepsin S with an N-terminal  
 257 Methionine-Proline and a C-terminal Glycine-6x-histidine tag  
 260 <220> FEATURE:  
 261 <221> NAME/KEY: MISC\_FEATURE  
 262 <222> LOCATION: (1)..(2)  
 263 <223> OTHER INFORMATION: Amino acid sequence for an N-terminal Methionine-Proline tag  
 265 <220> FEATURE:  
 266 <221> NAME/KEY: MISC\_FEATURE  
 267 <222> LOCATION: (3)..(333)  
 268 <223> OTHER INFORMATION: Amino acid sequence for Cathepsin S  
 270 <220> FEATURE:  
 271 <221> NAME/KEY: MISC\_FEATURE  
 272 <222> LOCATION: (334)..(340)  
 273 <223> OTHER INFORMATION: Amino acid sequence for a C-terminal Glycine-6x-histidine

tag

275 <400> SEQUENCE: 4  
 277 Met Pro Met Lys Arg Leu Val Cys Val Leu Leu Val Cys Ser Ser Ala  
 278 1 5 10 15  
 281 Val Ala Gln Leu His Lys Asp Pro Thr Leu Asp His His Trp His Leu  
 282 20 25 30  
 285 Trp Lys Lys Thr Tyr Gly Lys Gln Tyr Lys Glu Lys Asn Glu Glu Ala  
 286 35 40 45  
 289 Val Arg Arg Leu Ile Trp Glu Lys Asn Leu Lys Phe Val Met Leu His  
 290 50 55 60  
 293 Asn Leu Glu His Ser Met Gly Met His Ser Tyr Asp Leu Gly Met Asn  
 294 65 70 75 80  
 297 His Leu Gly Asp Met Thr Ser Glu Glu Val Met Ser Leu Met Ser Ser  
 298 85 90 95  
 301 Leu Arg Val Pro Ser Gln Trp Gln Arg Asn Ile Thr Tyr Lys Ser Asn  
 302 100 105 110  
 305 Pro Asn Arg Ile Leu Pro Asp Ser Val Asp Trp Arg Glu Lys Gly Cys  
 306 115 120 125  
 309 Val Thr Glu Val Lys Tyr Gln Gly Ser Cys Gly Ala Cys Trp Ala Phe  
 310 130 135 140  
 313 Ser Ala Val Gly Ala Leu Glu Ala Gln Leu Lys Leu Lys Thr Gly Lys  
 314 145 150 155 160  
 317 Leu Val Ser Leu Ser Ala Gln Asn Leu Val Asp Cys Ser Thr Glu Lys  
 318 165 170 175  
 321 Tyr Gly Asn Lys Gly Cys Asn Gly Gly Phe Met Thr Thr Ala Phe Gln  
 322 180 185 190  
 325 Tyr Ile Ile Asp Asn Lys Gly Ile Asp Ser Asp Ala Ser Tyr Pro Tyr  
 326 195 200 205  
 329 Lys Ala Met Asp Gln Lys Cys Gln Tyr Asp Ser Lys Tyr Arg Ala Ala  
 330 210 215 220  
 333 Thr Cys Ser Lys Tyr Thr Glu Leu Pro Tyr Gly Arg Glu Asp Val Leu  
 334 225 230 235 240  
 337 Lys Glu Ala Val Ala Asn Lys Gly Pro Val Ser Val Gly Val Asp Ala  
 338 245 250 255  
 341 Arg His Pro Ser Phe Phe Leu Tyr Arg Ser Gly Val Tyr Tyr Glu Pro  
 342 260 265 270

## RAW SEQUENCE LISTING

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:06

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw

```

345 Ser Cys Thr Gln Asn Val Asn His Gly Val Leu Val Val Gly Tyr Gly
346          275                      280                      285
349 Asp Leu Asn Gly Lys Glu Tyr Trp Leu Val Lys Asn Ser Trp Gly His
350      290                      295                      300
353 Asn Phe Gly Glu Glu Gly Tyr Ile Arg Met Ala Arg Asn Lys Gly Asn
354 305                      310                      315                      320
357 His Cys Gly Ile Ala Ser Phe Pro Ser Tyr Pro Glu Ile Gly His His
358                      325                      330                      335
361 His His His His
362                      340

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 07/17/2006  
PATENT APPLICATION:   US/10/646,470A      TIME: 09:32:07

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt  
Output Set: N:\CRF4\07172006\J646470A.raw

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,  
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:3,4

VERIFICATION SUMMARY

DATE: 07/17/2006

PATENT APPLICATION: US/10/646,470A

TIME: 09:32:07

Input Set : F:\CATS-5002-C1 Substitute Seq List.ST25.txt

Output Set: N:\CRF4\07172006\J646470A.raw